

## CLAIMS:

What is claimed is:

1. A method for normalizing two or more encoded text data strings, said encoded text data strings being expressed as a series of characters and meta data fields, said meta data fields containing all information regarding higher order control, formatting and display for the characters, said method comprising the steps of:
- receiving at least two strings for comparison;
- determining whether both or all of the strings are in similar or different meta normal form;
- responsive to determining that said strings are in different meta normal forms, converting one or more strings to a selected meta normal form such that all strings are in similar meta normal forms;
- comparing said strings to each other on a character-by-character basis irrespective of meta data fields in each string; and
- determining that said strings are equivalent if said step of comparing yields a match.
2. The method as set forth in Claim 1 wherein said step of converting comprises converting at least one string to Normal Form Meta Decomposed form.
3. The method as set forth in Claim 1 wherein said step of converting comprises converting at least one string to Normal Form Meta Composed form.
4. The method as set forth in Claim 1 wherein said step of comparing comprises: removing from said strings all meta data fields to yield pure strings

which are free of glyph and display-related information; and

comparing said pure strings on a character-by-character basis to determine the equivalency of said strings.

5. A computer readable medium encoded with software for

5 normalizing two or more text data strings, said text data strings being expressed as a series of characters and meta data fields, said meta data fields containing all information regarding higher order control, formatting and display for the characters, said software causing a computer to perform the following actions:

10 receive at least two strings for comparison;  
determine whether all of the strings are in similar or different meta normal form;  
responsive to determining that said strings are in different meta normal forms, convert one or more strings to a selected meta normal form such  
15 that all strings are in similar meta normal forms;  
compare said strings to each other on a character-by-character basis irrespective of meta data in each string; and  
determine that said strings are equivalent if said step of comparing yields a match.

20 6. The computer readable medium as set forth in Claim 5 wherein said software for converting one or more strings to a similar meta normal form comprises software for converting at least one string to Normal Form Meta Decomposed

form.

7. The computer readable medium as set forth in Claim 5 wherein said software for converting one or more strings to a similar meta normal form comprises software for converting at least one string to Normal Form Meta Composed form.

8. The computer readable medium as set forth in Claim 5 wherein said software for comparing said strings comprises software for performing the following actions:

removing from said strings all meta data to yield pure strings which are free of glyph and display-related information; and

comparing said pure strings on a character-by-character basis to determine the equivalency of said strings.

9. A system for normalizing two or more encoded text data strings, said encoded text data strings being expressed as a series of characters and meta data fields, said meta data fields containing all information regarding higher order control, formatting and display for the characters, said system comprising:

a meta form comparator for determining whether both or all of the strings are in similar or different meta normal form;

a meta form converter adapted to convert one or more strings to a selected meta normal form such that all strings are in similar meta normal forms responsive to a determination by said meta form comparator that said strings are in different meta normal forms;

a string content comparator for comparing said strings to each other on  
a character-by-character basis irrespective of meta data fields in each string;  
and

an equivalency evaluator for signaling that said strings are equivalent  
5 if said string content comparator indicates a character-by-character match is  
found.

10. The system as set forth in Claim 9 wherein said meta form converter  
is adapted to convert a string to Normal Form Meta Decomposed  
form.

10 11. The system as set forth in Claim 9 wherein said meta form converter is  
adapted to convert a string to Normal Form Meta Composed form.

12. The system as set forth in Claim 9 wherein said string content comparator  
comprises a meta data field filter for removing from said strings all meta data  
fields to yield pure strings which are free of glyph and display-related  
15 information such that said pure strings may be compared on a  
character-by-character basis to determine the equivalency of said strings.